CLAIMS

What is claimed is:

1. A system that facilitates decision tree learning, comprising:

a learning component that generates non-standardized data that relates to a split in a decision tree; and

a scoring component that scores the split as if the non-standardized data at a subset of leaves of the decision tree had been shifted and/or scaled.

2. The system of claim 1, further comprising a modification component that for a respective candidate split score, the data is modified by shifting and/or scaling the data and a new score is computed on the modified data.

3. The system of claim 1, further comprising an optimization component that analyzes the data and decides to treat the data as if it was: (1) shifted, (2) scaled, or (3) shifted and scaled.

4. The system of claim 1, the scoring component is employed for evaluating a data mining application.

5. The system of claim 1, the learning component processes continuous variable data or data subsets.

6. The system of claim 1, the scoring component generates evaluation indicating how well a model predicts continuous target data and whether or not the model is a suitable predictor for the target data.

7. The system of claim 6, the evaluation data is employed by users and/or subsequent automated components when determining model performance and/or selecting between models or model subsets.

8. The system of claim 1, the scoring component includes at least one of a data sample processor, a scoring constant, a gamma function, a matrix value, a vector value, and a mean value for data or a data subset.

9. The system of claim 1, the scoring component computes a Bayesian linear regression score as:

$$score = \pi^{-n/2}\left(\frac{\nu}{\nu+n}\right)^{1/2}\frac{\Gamma\left(\frac{\alpha+n}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}\left(\beta^{\frac{\alpha+r}{2}}\right)\frac{\left(\left|\mathbf{T}_n^{TR}\right|\right)^{-\left(\frac{\alpha+n}{2}\right)}}{\left(\left|\mathbf{T}_n^{R}\right|\right)^{-\left(\frac{\alpha-1+n}{2}\right)}},$$

$$\mathbf{T}_n = \mathbf{T}_0 + \mathbf{S}_n + \mathbf{U}_n$$
$$\mathbf{U}_n = \frac{\nu n}{\nu+n}(\bar{\mu}_0 - \bar{m}_n)(\bar{\mu}_0 - \bar{m}_n)'$$
$$\mathbf{S}_n = \sum_{i=1}^{n}(\bar{x}_i - \bar{m}_n)(\bar{x}_i - \bar{m}_n)'$$
$$\bar{m}_n = \frac{1}{n}\sum_{i=1}^{n}\bar{x}_i$$

wherein bold-face symbols denote square matrices, symbols with overlines denote (one dimensional) vectors, the ' symbol denotes transpose, and $|\ |$ denotes determinant, n represents a number of records in the data, Γ is a gamma function satisfying $\Gamma(x) = (x-1)\,\Gamma(x-1)$, $\bar{x}_i$ denotes a vector of values for relevant variables in an *ith* case in the data, the superscripts TR and R in $\mathbf{T}_n^{TR}$ and $\mathbf{T}_n^{R}$ denote that the matrices are defined with respect to target and regressor variables in a first case and regressor variables in a second case.

10. A computer readable medium having computer readable instructions stored thereon for implementing the scoring component of claim 1.

11. A system that facilitates data mining, comprising:

means for automatically generating a set of non-standardized data associated with a set or subset of data relating to a continuous variable, the non-standardized data associated with a split in a decision tree; and

means for automatically scoring the split as if the non-standardized data were shifted and/or scaled.

12. The system of claim 11, further comprising means for determining whether to perform the shifting and/or scaling operations.

13. The system of claim 11, further comprising means for shifting and/or scaling the set or subset of data relating to the continuous variable.

14. A method that facilitates decision tree learning, comprising:

determining whether to perform a virtual shifting and/or scaling operation on a non-standardized set of data associated with leaves of a decision tree; and

automatically assigning scores to the leaves based in part upon the determination of whether to perform the virtual shifting and/or scaling operation.

15. The method of claim 14, further comprising performing at least one actual scaling and/or shifting operation on the non-standardized set of data.

16. The method of claim 14, further comprising processing a model in a form of a linear regression.

17. The method of claim 14, the virtual shifting operation includes omitting a matrix operation from the assignment of scores.

18. The method of claim 14, the virtual shifting operation includes modifying a subset of elements relating to a covariance matrix.

19. The method of claim 14, determining at least one constant value before assigning the scores.

20. The method of claim 19, the constant value relates to diagonal elements of a matrix and is assigned a value of about 0.01.

21. A computer readable medium having a data structure stored thereon, comprising:

a first data field describing a non-standardized set or subset of data relating to a continuous variable;

a second data field describing a decision tree and associated branches; and

a third data field describing a score for the branches, the score computed for the branches as if the non-standardized set of subset of data had been shifted of scaled.

22. The computer readable medium of claim 21, further comprising a data field to indicate at least one of a virtual shifting operation and a virtual scaling operation.

23. The computer readable medium of claim 21, further comprising a data field to indicate at least a portion of the non-standardized set or subset of data is to be shifted and/or scaled.

24. A data packet that passes between at least two computer processes, comprising:

a first data field describing a non-standardized set or subset of data relating to a continuous variable;

a second data field describing a decision tree and associated branches; and

a third data field describing a score for the branches, the score computed for the branches as if the non-standardized set of subset of data had been shifted and/or scaled.